

Proceedings of ISON 2013, 4<sup>th</sup> Interactive Sonification Workshop, Fraunhofer IIS, Erlangen, Germany, December 10, 2013

# INTERACTIVE SONIFICATION TO SUPPORT JOINT ATTENTION IN AUGMENTED REALITY-BASED COOPERATION

Alexander Neumann, Thomas Hermann, René Tünnermann

Ambient Intelligence Group  
CITEC - Bielefeld University  
Bielefeld, Germany

[alneuman,thermann,rtuenner]@techfak.uni-bielefeld.de

## ABSTRACT

This paper presents and evaluates interactive sonifications to support periphery sensing and joint attention in situations with a limited field of view. Particularly Head-mounted AR displays limit the field of view and thus cause users to miss relevant activities of their interaction partner, such as object interactions or deictic references that normally would be effective to establish joint attention. We give some insight into the differences between face-to-face interaction and interaction via the AR system and introduce five different interactive sonifications which make object manipulations of interaction partners audible by sonifications that convey information about the kind of activity. Finally we present the evaluation of our designs in a study where participants observe an interaction episode and rate features of the sonification in questionnaires. We conclude the results into factors for acceptable sonifications to support dyadic interaction.

## 1. INTRODUCTION

In natural human-human interaction, we command over many communicative resources to coordinate joint activity, such as speech, gaze, deictic gestures or head gestures. Their interplay allows us to establish and sustain joint attention when needed, such as in collaborative planning tasks. We deal with the latter in an interdisciplinary project between linguistics and computer science where we aim for a better understanding of the principles of successful communication<sup>1</sup>. We have introduced and developed an Augmented Reality (AR) system that enables us to ‘(de-)couple’ two users interacting co-presently at a table in a cooperative task of planning a recreational area. The AR system allows us to precisely record what the interaction partners see at any moment in time – and thus to understand their next actions based on the information they have selected. Besides the capability of visual interception, we extended the system to also enable an auditory interception by using microphones and in-ear headphones.

Yet we can also *manipulate* the media (both visual and auditory cues) in manifold ways: first by introducing *disturbances* to study how these are compensated in interaction, and secondly, by *enhancements*, to contribute to wearable assistance systems that better support cooperating users.

We have proposed and introduced various new sonic enhancement methods in [8] to increase the users’ awareness of their interaction partner. In [3] we used Conversation Analysis of a multimodal corpora of interacting users to identify which cues are rel-

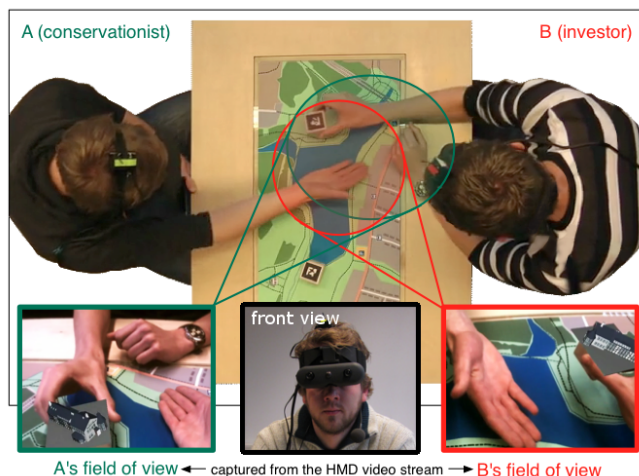


Figure 1: Two participants argue about the future of an area around the Bielefelder Obersee. Video streams from their *Head-Mounted Displays* (HMDs) are analyzed and processed in real-time. The markers on top of the wooden cubes are augmented with models representing concepts for possible projects (e.g. *hotel*).

evant for establishing and maintaining joint attention and to find specific problematic occasions which could be solved by such a method. In this paper, we take the next step and evaluate the approaches at hand of a user study with test listeners. The aim is to better understand the principles of how sound can be successfully used, and what sounds are accepted. We continue with a brief summary of our project, hardware setup and basic task.

## 2. ALIGNMENT IN AR-BASED COOPERATION

In the Collaborative Research Center 673 *Alignment in Communication* we combine proven communication research methods with new interdisciplinary approaches to get a better understanding of what makes communication successful and to gather insights into how to improve human-computer interaction. The C5 project *Alignment in AR-based cooperation* uses emerging Augmented Reality technologies as a method to investigate communication patterns and phenomena. In experiments we ask users to solve tasks collaboratively, using an Augmented Reality based Interception Interface (*ARbInI*) which consists of several sensors and displays and

<sup>1</sup>[www.sfb673.org/projects/C5](http://www.sfb673.org/projects/C5)

allows us to record and alter the perceived audiovisual signals of a system's users in real-time. For data analysis we combine the benefits of machine-driven quantitative data mining approaches with qualitative conversation analysis in a mutual hypothesis generation and validation loop.

### 2.1. Obersee Scenario

Our current experimental task is a fictional recreation scenario of the surroundings of the Bielefelder Obersee, the largest lake in Bielefeld. The main idea is to let two opposing parties argue about the future shape of this area. The participants are seated at a table with a map of this area, equipped with symbolic representations of possible attractions or construction projects as shown in Figure 1. These 'symbolic representations' are wooden cubes with ARToolkitPlus markers on top of them. To elicit some initial 'disagreement' we ask the participants to argue from the contrary points of view of an 'investor' interested in attracting many tourists and a 'conservationist' aiming at the preservation of wildlife. Both parties have to overcome their opposing goals and agree on a final result which should be presented after 20 minutes of negotiation.

When participants look at a cube through their *Head-Mounted Displays* (HMDs), the system detects the marker and augments a virtual representation of the attraction previously connected to this marker at the spot where the marker was detected. Object size and orientation varies according to the marker's position within the participant's field of view. This feature allows us to monitor, control and manipulate the visual information available to both users separately during the negotiation process at every moment during the experiment [1].

## 3. MUTUAL MONITORING IN FACE-TO-FACE AND AR-BASED INTERACTION

In natural face-to-face interaction, participants rely on the possibility of mutual monitoring and on-line analysis of the co-participant's actions (speech, bodily conduct, gesture etc.). This enables them to adjust their ongoing actions on a fine-grained level to each other.

A conversational analysis of interactions in the described setup has shown several emerging problems due to the used augmented reality gear [3]. In summary these are:

- Mutual monitoring-based procedures enable interlocutors to prevent emerging parallel activities. This ensures the sequential organization of their activities.
- The lack of mutual monitoring in AR leads to cases where both participants initiate actions simultaneously without a mechanism to repair the situation quickly, as would be the case in face to face conversation.
- There is only a short period of time to repair emerging parallel activities.

The lack of mutual monitoring requires a mechanism to compensate this lack of mutual awareness. The compensation has to be done within a short time window of a few seconds, in order to prevent simultaneous actions by the actors. Since the field of view is limited - which is common in augmented reality systems [7, 11] - and visual augmentations would eventually lead to time-intensive search processes, sound is an attractive and neglected channel. The following section will approach and develop this idea from an ecological listening perspective.

## 4. AUDITORY DISPLAYS FOR NON-VISUAL GUIDANCE OF ATTENTION

In everyday interaction sound is an important cue to catch and orient our focus of attention, as for instance exemplified by situations where we hear our name being called from somewhere, a sudden explosion or an approaching car on the street [6]. However, there are also many situations where not a sudden event, but a change of sound draws our attention even if it is only subtle. For instance when driving a car and suddenly experiencing a change of the engine sound. These examples demonstrate how sound is effective for the organization of our attention in natural situations. Certainly this can also be transferred to technical systems: the Geiger counter is a device that represents radiation by a granular sonic texture, drawing attention as the rate changes; the pulse oximeter device is indispensable to auditory monitor heart rate and oxygen level in blood during surgeries.

*Sonification* enables us to profit from our auditory information processing which operates largely in parallel and independently of our primary task. For instance, in [5], we have presented a sonification of sport aerobics movements which enables the listeners to understand various features of their exercise, e.g. how fast the movement is executed and when the exercise changes. The system was primarily targeted at visually impaired users to improve their participation in aerobics. Another recent sonification system, which we developed in context of and for our AR-system is the sonification of head gestures such as nodding and shaking the head: as the head-mounted displays allow either to look on the desk or to look to the interaction partner, but not simultaneously, the sonification of head gestures conveys analogical and subtle information to support interaction [4]. Furthermore, enhancing and augmenting object sounds with informative or aesthetic acoustic additions is a well established approach in Sonic Interaction Design [9], yet so far rarely considered for collaborative applications [2].

With this motivation and context, we now summarize our most recent development, the sonification of object interactions for supporting dyadic interaction which we introduced as idea and method in [8]. Manipulations of our physical environment usually produce feedback sounds on what, where and how strong we interacted. As the sounds propagate not only to our own ears, but also to others in the surrounding, they can be used to become and stay aware of activities in the environment. An office worker for instance could know without looking, if her colleague is typing or not, only from the existence or absence of interaction sounds with the keyboard. Features such as writing speed, error rate and perhaps even the urgency of the writing may be picked up as well. Parents often use sound as a display for their children's activities out of their sight. Here, actually, the absence of steady noises is an important cue that something might not be right and thus needs attention.

Sound draws our attention towards events outside our field of view, e.g. somebody approaching from behind, or a mobile phone beeping on the table [10]. We make use of this specific capacity of sound for AR-based cooperation to create an awareness of events happening outside the typically very limited view angle of head-mounted displays. We argue that listeners are well capable to interpret physical interactions correctly from interaction sounds, and thus they draw subconsciously conclusions about the source of a heard sound. From that motivation we developed a set of sonification methods, that not only imitate (and exaggerate) natural physical interactions, but also allow to associate sounds to nor-

mally silent actions such as carrying objects through air. From these methods we selected five for the following study, which will be explained in the following section.

## 5. SONIFICATION DESIGNS

We are mainly interested in the object interactions (a) to move (shift/rotate) it on the desk, (b) to pick/lift an object, (c) to carry it to a different location through air, and finally (d) to place it on the desk.

Such interactions are ubiquitous in our scenario and are partly accompanied naturally with interaction sounds (in our scenario: of wooden objects touching our glass table), specifically only (a), (b) and (d). Some actual interactions are silent (e.g. c), and many interaction go unnoticed as they can be and often are executed rather silently. The artificial sonification of the interaction types are meant to reliably make the interaction partners aware of these activities.

The data used to practically implement our sonifications were captured by a downwards looking camera mounted on the ceiling and tracked with ARToolkit. The derivation of 'high-level' features that correspond to our interaction classes (a–d) is a complex computational process which is beyond the scope of this paper, but works reliably enough to provide the basis for the sonifications. The feature extraction results in either continuous features such as the current velocity, position or rotation of an object, or discrete events such as lifting or putting objects. With these tracking data we implemented five sonifications, namely Direct Parameter Mapping (PM), Abstract signals (AS), Exaggerated samples (ES), Naturalistic imitation (NI), object-specific sonic symbols (OS), which we explain next. A brief overview is also shown in Table 1. Example videos with overlaid sonification are available at our website<sup>2</sup>.

### 5.1. Direct Parameter-Mapping Sonification

In this method we rather directly turn the multivariate times series of features into sound. We use time-variant oscillators with frequency and amplitude parameters and map the vertical height of an object above the table to frequency, following the dominant polarity association [12]. The frequency range is 100Hz to 300Hz using sine tones without higher harmonics, so that the resulting sound is both rather quiet and has limited interference with the concurrent verbal engagement of the users. This approach is rather disturbing as objects create sine sounds all the time. We have also created a version that controls the amplitude from the current object velocity but such an excitatory mapping was not selected for this study.

### 5.2. Abstract Signal Sonification

This design signals events by clear and distinguishable abstract sounds:

- Lifting is represented by a short up-chirped tone.
- Putting an object down leads to a down-chirped tone.
- Pushing an object on the desk surface is sonified by pink noise that decays smoothly after the action stops.
- Carrying an object above the surface leads to low-pass filtered white noise, again with smoothly decaying level as the action stops.

<sup>2</sup><http://www.techfak.uni-bielefeld.de/ags/ami/publications/NHT2013-ISS/>

The sounds may be understood as abstractions of sand and wind sounds for translation on ground or in air.

### 5.3. Exaggerated Samples

This sonification design is similar to the Abstract Signal sonification, yet we here used more obtrusive sounds, to examine how they cause problems or disturb ongoing interaction. For the actions 'lift', 'put', 'pushing' and 'carrying' we chose a high pitched blings for lift, crashing windows for put, creaking for pushing an object and a helicopter for carrying, in order to render the actions very salient.

### 5.4. Naturalistic Imitation

Assuming that naturalistic sounds will be most easily understood, we created a sonification that uses the familiar sound bindings as true as possible. However, our sonification is different from what would be obtained by attaching a contact microphone to the table and amplifying the real sound signals in (a) that even silently executed actions (such as putting an object on the table) here leads to a clearly audible put-sound, and (b) that we here gain the conceptual ability to refine the sounds (as parameterized auditory icons) dependent on actions and circumstances we regard as important. We could for instance control the level or brilliance of a sound by how far the object is outside the interaction partner's view. The samples used have been recorded using a microphone and the same wooden objects that are used in the AR scenario.

### 5.5. Object-specific sonic symbols

Finally, we selected the sound to correspond to the model being shown on top of our objects. For instance while manipulating the 'playground' placeholder object, a sample recorded on a playground is played. Likewise for the petting zoo, animal sounds evoke the correct association. Technically, sample playback is activated whenever (but only if) an object is moved around, ignoring the object's height above the desk. The sound is furthermore enriched by mapping movement speed to amplitude and azimuthal position to stereo panning, creating a coarse sense of directional cues.

## 6. EVALUATION

To examine how the sonifications are understood by listeners and how they might affect interaction, we conducted a pilot study, asking subjects to rate the different sonifications of a given interaction example according to a number of given statements. We focused on three research questions:

- How do the sonifications perform concerning interaction with speech, obtrusiveness, utility, aesthetics, learnability and distinguishability?
- Which designs perform better; which perform worse and why?
- Is there a clear winner? If not: How do the most promising designs differ?

Table 1: The five presented prototypes vary in representation and represented features. While parameter mapping (PM) uses analogue sounds to represent height above ground and movement speed, Abstract Signals use more symbolic sounds to signalize four discrete events. Object-specific sounds (OS) indicate only activity and location with a sample semantically connected to the handled object.

|          | PM                    | AS                        | ES                              | NI                 | OS |
|----------|-----------------------|---------------------------|---------------------------------|--------------------|----|
| Category | Parameter Mapping     | Earcon                    | Auditory Icon                   |                    |    |
| Sounds   | synthesized/generated |                           | samples (recorded, synthesized) | samples (recorded) |    |
| Opout    | continous             | discrete                  |                                 |                    |    |
| Features | velocity, location    |                           |                                 |                    |    |
|          | height above ground   | lifting, putting, pushing |                                 |                    |    |
|          |                       | carrying                  |                                 |                    |    |

|    |                               |
|----|-------------------------------|
| PM | Parameter Mapping             |
| AS | Abstract Signals              |
| ES | Exaggerated Samples           |
| NI | Natural Imitation             |
| OS | Object-specific Sonic Symbols |

### 6.1. Study Design

A short video clip showing a real dyadic interaction of the Obersee scenario from the top perspective was augmented with the different sonification approaches as explained before. The interactions shown in the video were thereby directly coupled with the sonifications.

The resulting five audio-visual stimuli were randomized for each participant in this within-subject design. Each participant first received an introduction and the opportunity to look at the interaction before the main experiment started. Participants were asked to watch the video (several times, if they like) until they had a good idea what goes on to fill a questionnaire for the stimulus. The questionnaire contained statements and questions, and a 7-point Likert scale ranging from 1 ('false') to 7 ('true') (resp. 'no' to 'yes'). The questions/statements to be answered for each method are listed in an English translation in Table 2. Additionally, we included a free text field to collect suggestions and ideas for each design. We also collected basic data such as age, sex and profession as well as information about experience with computers and musical instruments and possible issues related to sound awareness.

### 6.2. Results

We interviewed 23 participants (15 male) between 20 and 33 (average 27.5). Most of the participants were students from various disciplines. The variance analysis for every question was done with an ANOVA with a threshold significance level of  $p_a < 0.01$ . Out of the 22 questions 3 (A14, A18, A20) questions do not fulfill this criteria. However,  $p_a$  values for A14 and A20 are only slightly higher ( $p_a = 0.012$ ) and can be considered significant with a level of significance of  $p_a = 0.05$  which still is an acceptable choice in our scenario. To identify differences and trends we used standard t-tests as a significance measure. When we state in the following that an approach is better or worse than the others this means that an independent two-samples t-test revealed significant difference

Table 2: Method specific statements from our questionnaire.

| ID  | question   |
|-----|--|
| A1  | I can well follow the dialogue   |
| A2  | I can perceive and distinguish the sounds even when I attend to the speech           |
| A3  | I attend mainly to the sounds  |
| A4  | Sounds cover language and are thus distracting                                       |
| A5  | Dialog is central for me, even when I perceive the sounds well                       |
| A6  | Interaction sounds are informative   |
| A7  | Interaction sounds are obtrusive   |
| A8  | Interaction sounds are pleasant  |
| A9  | Interaction sounds are comprehensible  |
| A10 | Interaction sounds are disturbing  |
| A11 | Interaction sounds are well-sounding   |
| A12 | Interaction sounds are irritating  |
| A13 | Interaction sounds are distracting   |
| A14 | I got used to the sounds on several listening  |
| A15 | I can imagine to use the sounds for extended time, if they would improve cooperation |
| A16 | The interplay of individual sounds is well   |
| A17 | I can associate the sounds with a metaphor that explains the sounds                  |
| A18 | I need to learn-by-heart the meaning of sounds                                       |
| A19 | The object-put sound is well done  |
| A20 | The object-lift sound is well done   |
| A21 | The object-on-desk-shift sound is well done  |
| A22 | The object-carrying-sound is well done   |

between two samples where the first sample contains the results for the approach and the second sample the results of all the other approaches. Positive or negative tendencies were identified with a one-sampled t-test which we used to test if the results of one approach differs significantly from the neutral rating 4. If not mentioned otherwise the level of significance  $\alpha$  is 1%.

### 6.3. The Interplay of Sounds and Dialog

An important aspect is how object interaction sounds work together with the ongoing verbal interaction, particularly as the sonifications are intended to augment the cooperative planning that involves intensive verbal negotiations. The first block of questions/statements aims at elucidating the interplay of sonifications and verbal sounds. Results are depicted in Figure 2.

In result, all sonifications allow still to follow the dialog (A1) with Natural Imitation (NI) performing significantly better and

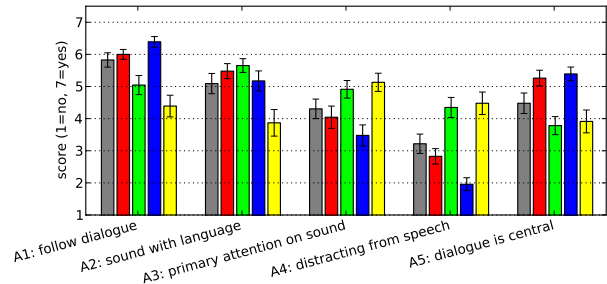


Figure 2: Average scores and standard error by questions and methods (PM (gray), AS (red), ES (green), NI (blue), OS (yellow)). NI was rated the least distracting design which did not interfere with the dialog.

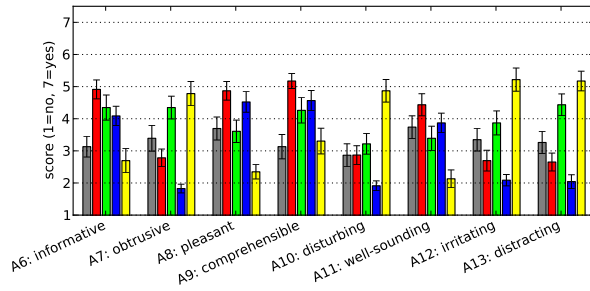


Figure 3: Average scores and standard error by questions and methods (PM (gray), AS (red), ES (green), NI (blue), OS (yellow)). AS was perceived as the most useful design while OS got the least favorable scores.

the Object-specific Sonic Symbols (OS) performing significantly worse than the other designs. Concerning the compatibility of sound and speech (A2) only OS performed lower than neutral. OS was also perceived as the most present design (A3) and also as distracting (A4) together with the Exaggerated Samples (ES). In contrast, NI was rated as the least obtrusive design and the least distracting approach. This leads to the expectable result that NI was rated the approach leading to an experience where the dialog was central (A5). ES performed worst here. It is noteworthy that OS performs equally or even worse than ES even though ES was designed to cover language while OS was meant to be ambient.

#### 6.4. Influences of Sound on the User

In the next group of questions (A6–A13) we were eager to learn how the different sonifications compare in terms of qualitative effects, as shown in Figure 3.

As expected, the naturalistic imitations are the least obtrusive (A7), least disturbing (A10), least irritating (A12) and least distracting (A13). The reason for that might be the fact that there are less sounds played in this sonification: carrying an object in air is silent and thus not represented by sound. Apart from this exception, the ratings can well be regarded as a baseline to which the other methods need to be compared to.

An unexpected counterpoint is the very obvious bad result of the object-specific sonification method: It is the least informative (A6), most distracting, irritating, disturbing and obtrusive approach. It is also rated the least pleasant (A8) and worst-sounding (A11) design. In contrast, only the Abstract Signals (AS) was rated as rather well-sounding and achieved a score better than neutral. Additionally, AS was perceived as the most informative (A6), pleasant and comprehensible (A9) choice. The participant also found the parameter mapping to be difficult to grasp and rated it the least comprehensible.

#### 6.5. Temporal Aspects and Understanding

Let us look on how the sonifications are rated concerning the long-term usability, shown in Figure 4. Certainly, participants can only vaguely extrapolate from their short experience. For instance, we cannot say anything about learnability (A18) since no significances could be found due to the high variance of the given answers.

However, some conclusion can be drawn. The participants anticipated that getting used to AS would be most likely (A14) and

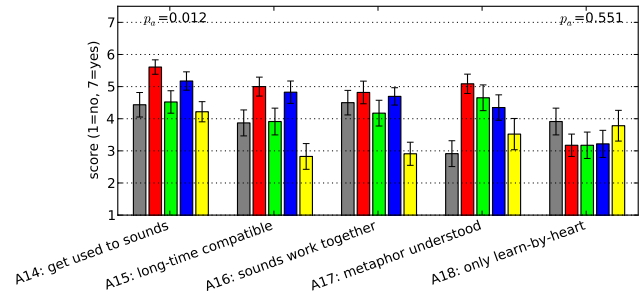


Figure 4: Average scores and standard error by questions and methods (PM (gray), AS (red), ES (green), NI (blue), OS (yellow)). The participants stated that they could get used to AS and felt like they understood the underlying metaphor. Again, OS scored significantly lowest in most categories. The ANOVA results also indicate no significant differences for A18. A14 also misses the threshold of  $p_a < 0.01$  slightly.

cast doubt on the long-term compatibility of OS (A15). One reason might be that they found the object specific sound not working very well together (A16). The Parameter Mapping (PM) fails to transport the underlying metaphor (A17) and receives the lowest score which probably explains the lack of comprehensibility mentioned earlier. Abstract sounds and their meaning were mostly understood and also the only approach that scores above neutral.

#### 6.6. Relation of Event and Sounds

In the final part of the questionnaire we asked about the distinguishably/recognizability of the sonified events. The results are shown in Figure 5.

AS was favored concerning shift (A21) and carry (A22) sounds and also is the only design which scores better than neutral for shifting and slightly above neutral (level of significance of  $\alpha = 5\%$ ) for lifting and carrying. Participants also rated object placing sonifications (A19) of AS and NI positively.

This time it is no surprise that OS scores the lowest in all of the mentioned categories since this method does not distinguish between events as mentioned in Section 5. The real surprise here is a

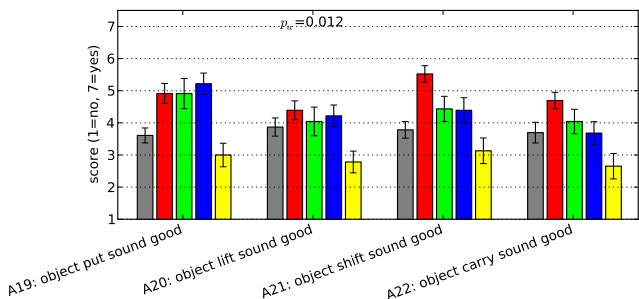


Figure 5: Average scores and standard error by questions and methods (PM (gray), AS (red), ES (green), NI (blue), OS (yellow)). While there was no clear favorite for the put and lift event, the participants preferred AS for carrying and shifting. A20 misses the ANOVA threshold of  $p_a < 0.01$  slightly.



score above zero for NI for its carrying sound since there was none. In our initial theory we assumed that to deal with this question not making sense some participants chose the ‘neutral element’ (score 4) while some others went for the lowest score. However, since some participant rated the NI’s carrying sonification (silence) as ‘very good’ and ‘good’ this theory was rejected.

In sum we observed that in most cases either NI or AS were rated best while OS usually scored worst or similar to ES which was a surprise for us. The fact that the parameter mapping was not understood by most of the participant might explain its result of being never favored but also never fell back behind the other approaches.

### 6.7. Similarity in the Evaluation Space

As mentioned above we used the independent two-sample t-test to identify the best and worst performing approach in every category. However, observations of the results also show that some designs score in a similar way which is ignored by this 1-vs-4 sample splitting. To measure similarity we treat every set of answers as a 23-dimensional vector and calculate the angle between the two answer vectors which is a common practice in text mining, especially in combination with the bag-of-words model. A small angle indicates similarity. The comparison of the five mean vectors and the angles between them can be seen in Figure 6.

As expected AS and NI are indeed relatively similar. With these findings in mind we considered AS and NI sharing a subspace of the whole evaluation space and repeated the independent two-sample t-test with a 2-vs-3 sample split. As a result, the coupled NI/AS performance was always at least equal but most of the time better than the PM/ES/OS performance. In consequence of this, we consider NI and AS as the most promising sonification prototypes presented in this study.

## 7. DISCUSSION

The results of our study show clear implications on the basis of 23 subjects rating statements and answering questions for all the 5 sonification methods. AS and NI both were perceived positively regarding most of the investigated categories. Their characteristic differences make them suitable for slightly differing fields of application. In cases where movement sonification should be a prominent feature AS should be favored since it was rated the most informative, pleasant and comprehensible design. In other scenarios where speech and verbal understanding must not be interfered by movement sonifications, we recommend NI since it was the least disturbing and least distracting approach. Both prototypes will be improved during the next design iteration. Especially overall aesthetics, event representation and long-term acceptance ratings imply potential for improvement.

As mentioned earlier the low performance of OS is surprising but there are some evidences which could explain the participants’ issues with this approach. First of all, people stated that they had problems understanding the metaphor behind this concept. We mentioned that the subjects in the video see a playground augmented on top of the wooden cube, but it was not visible in the video stream. The viewers only saw the wooden cube. This made it harder to connect the children’s sound to the playground. We assume that another object and therefore another sample could have led to better results. Ambient noises emitted by kindergartens and playgrounds are controversial and regarded by certain people

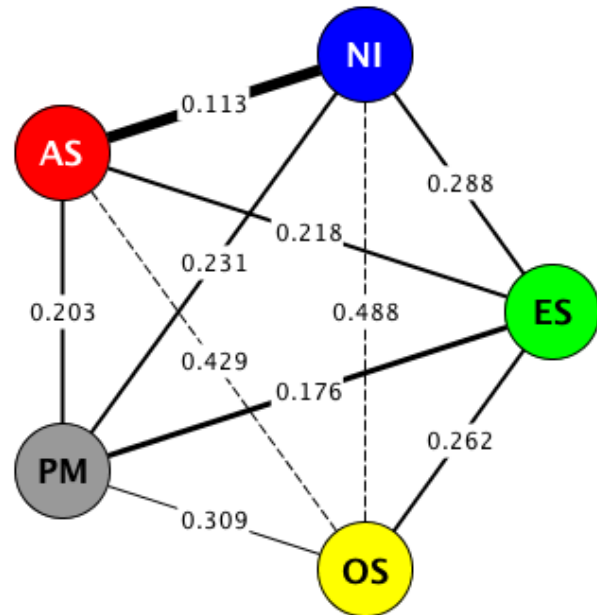


Figure 6: Calculating the angle between the 23-dimensional mean vectors of every prototype revealed interesting relations. A thicker line indicates higher similarity. The small angle between AS and NI support the impression that both approaches were rated similarly.

as distracting and disturbing. This might explain why even helicopter and crashing sounds used in ES were perceived as less disturbing.

The importance of an easy to understand metaphor is also indicated by the performance of the parameter mapping which was average at best. A clear connection between the movement and the sounds would probably lead to an improved experience since the chosen sounds did not vary much from AS which was rated significantly better.

An issue which influenced all sample based sonifications is the chopped sound caused by short movements (also discussed in [8]) which was perceived as unpleasant by most participants. NI is influenced less because in the chosen video sample most short movements happen in the air. A well chosen attack and decay time might reduce this issue but still allow to identify short movements.

Even though this study was suited to identify general characteristics for future movement sonifications, an interaction study has to follow to investigate the usability in an interactive scenario. Overlap was excluded but will frequently appear in the described field of application and will make it more difficult to identify the currently moved object(s). In these cases object specific sound characteristics could be helpful.

## 8. CONCLUSION

We have presented the results of a user study to evaluate five initial prototypes to support joint attention in dyadic augmented reality-based cooperation. These five sonification approaches were created to offer better awareness of the interaction partner’s object

manipulations, ranging from naturalistic over exaggerated and abstract sonifications to sounds that allow object identification.

In summary, the abstract sonification and the naturalistic imitation sonification were well perceived and rated positively. In situations where the information should stay in the background, naturalistic sonification is a good choice since it was rated as the least interfering design. In other scenarios where the information is of a major interest, abstract sonification is a better candidate since it was perceived as the most informative, pleasant and comprehensible approach. Also a blend between naturalistic and abstract sonification, using parameterized auditory icons would be an interesting candidate for further evaluation.

## 9. ACKNOWLEDGMENTS

This work has partially been supported by the Collaborative Research Center (SFB) 673 Alignment in Communication and the Center of Excellence for Cognitive Interaction Technology (CITEC). Both are funded by the German Research Foundation (DFG).

## 10. REFERENCES

- [1] A. Dierker, C. Mertes, T. Hermann, M. Hanheide, and G. Sagerer. Mediated attention with multimodal augmented reality. *Proceedings of the 2009 international conference on Multimodal interfaces - ICMI-MLMI '09*, page 245, 2009.
- [2] W. W. Gaver. Sound support for collaboration. In *Proceedings of the Second Conference on European Conference on Computer-Supported Cooperative Work, ECSCW'91*, pages 293–308, Norwell, MA, USA, 1991. Kluwer Academic Publishers.
- [3] T. Hermann, A. Neumann, C. Schnier, and K. Pitsch. Sonification for supporting joint attention in dyadic augmented reality-based cooperations. *AM '13: Proceedings of the 8th Audio Mostly Conference: A Conference on Interaction with Sound*, 2013.
- [4] T. Hermann, A. Neumann, and S. Zehe. Head gesture sonification for supporting social interaction. *Proceedings of the 7th Audio Mostly Conference on A Conference on Interaction with Sound - AM '12*, pages 82–89. ACM Press, 2012.
- [5] T. Hermann and S. Zehe. Sonified aerobics - interactive sonification of coordinated body movements. *The 17th Annual Conference on Auditory Display*, Budapest, Hungary 20–24 June, 2011, *Proceedings. OPAKFI*, 2011.
- [6] C. Ho and C. Spence. Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. *Journal of experimental psychology: Applied*, 11(3):157, 2005.
- [7] A. Janin, D. Mizell, and T. Caudell. Calibration of head-mounted displays for augmented reality applications. In *Virtual Reality Annual International Symposium, 1993., 1993 IEEE*, pages 246–255, 1993.
- [8] A. Neumann and T. Hermann. Interactive sonification of collaborative ar-based planning tasks for enhancing joint attention. *Proceedings of the 19th International Conference on Auditory Displays*, pages 49–55. *The International Community for Auditory Display (ICAD)*, 2013.
- [9] S. Serafin, K. Franinović, T. Hermann, G. Lemaitre, M. Rinott, and D. Rocchesso. Sonic interaction design. In T. Hermann, A. Hunt, and J. G. Neuhoff, editors, *The Sonification Handbook*, chapter 5, pages 87–110. Logos Publishing House, Berlin, Germany, 2011.
- [10] C. Spence, J. Ranson, and J. Driver. Cross-modal selective attention: on the difficulty of ignoring sounds at the locus of visual attention. *Perception & psychophysics*, 62(2):410–24, Feb. 2000.
- [11] A. Tang, C. Owen, F. Biocca, and W. Mou. Comparative effectiveness of augmented reality in object assembly. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 73–80. ACM, 2003.
- [12] B. N. Walker and G. Kramer. Mappings and metaphors in auditory displays. *ACM Transactions on Applied Perception*, 2(4):407–412, Oct. 2005.